CS412 Machine Learning - 2020

COURSE PROJECT

The purpose of the project is to increase your knowledge about machine learning and get hands-on practical experience.

**Project topic:**

* You are given a problem about **Kaggle data scientists** in the attached ppt. Originally it is a public dataset but we have simplified it.

The project consists of two tasks: a prediction problem (predict **job satisfaction level** of a data scientist, based on their level/knowledge/current job/…), as well as analysis of the data to draw some insights. More details can be found in the **DETAILS. ppt file** (about what is asked, evaluation, submission…) and the Kaggle site:

<https://www.kaggle.com/t/ae0b72b683954031a7a8404bea93e8a5>

* If you also have a graduation project that has a **substantial ML problem that you can describe in 2 lines as an ML problem,** you can email me that description and I will decide (deadline Dec 24).

“given ….. as features, we want to predict/find/cluster……”

**Groups**: You will work in groups of 1 to 4. Please indicate your groups in the Google sheet below:

<https://docs.google.com/spreadsheets/d/1ogdkYzY06D3qCE2kxnl4mSAkk3XQSqeztlGyMpEVIU0/edit?usp=sharing>

**What to submit:** A **ppt presentation** of your project **and a Kaggle submission**.

**Grading**: You will have a small **exam/interview** at the end about your project about stuff you have done in your project. Your grade will be based on the performance you achieved in your project (e.g. prediction accuracy) and your individual exam performance (80-20%).